Laura E. Cowley[a], Sabine A. Maguire[a], Daniel M. Farewell[a], Alison M. Kemp[a]

Affiliations:

[a]Division of Population Medicine, School of Medicine, Cardiff University, Cardiff, Wales, United Kingdom
CowleyLE@cardiff.ac.uk, sabinemaguire@gmail.com, FarewellD@cardiff.ac.uk, KempAM@cardiff.ac.uk

Address correspondence to: Laura E. Cowley, Division of Population Medicine**,** School of Medicine, Neuadd Meirionnydd, Heath Park, Cardiff University, CF14 4YS, United Kingdom. Telephone: 0044 2920 688688 E-mail: CowleyLE@cardiff.ac.uk

To the Editor,

We read with great interest the critical analysis of our work (Cuellar, 2017). We appreciate the opportunity to respond, and to provide clarification regarding the purpose of our validated Predicting Abusive Head Trauma (PredAHT) clinical prediction tool (Maguire et al., 2011; Cowley et al., 2015).

In the abstract and throughout the article, Cuellar has somewhat oversimplified the objective of PredAHT, stating that it was developed "to make an AHT diagnosis" (Cuellar, 2017 p. 223). We emphasise that PredAHT was not designed to be a diagnostic tool, but as an aid to "assist frontline professionals when deciding whether to refer a child for specialist clinical and multiagency investigation of possible AHT" (Maguire et al., 2011 p. e558). As with other clinical prediction tools, PredAHT must be used in combination with all clinical, forensic, historical and psychosocial information available in each individual case. It "will not confirm or exclude AHT in isolation" (Cowley et al., 2015 p. 296), and "will never replace the diagnostic skills of the clinician" (Maguire et al., 2011 p. e558). We are somewhat puzzled by the statement in the abstract, that PredAHT is "an inappropriate foundation upon which to base an opinion that will be used in a criminal prosecution". We would agree entirely, as we have never suggested that the purpose of this clinical tool is for use as the foundation or basis of an opinion for a criminal prosecution, but rather that it may

"assist clinicians offering medical testimony in civil or criminal proceedings, in demonstrating why certain combinations of features are more or less predictive of an abusive etiology" (Maguire et al., 2011 p.e558). This is a tool for clinicians to gauge the likelihood of AHT in a child with a given set of clinical features and not a tool designed for legal purposes *per se*; we fully acknowledge that medicine cannot do the job of the law. The role of the clinician in the UK, together with the children's social care team is to determine, on the balance of probability, whether the child has suffered from child abuse such that the family can be supported and the child and siblings can be protected from future harm. Clinical prediction tools, including PredAHT, are designed for use by clinicians when they are assessing patients with complex conditions. PredAHT can be used as a basis to explain to the courts how, at a certain point in the clinical assessment process, the different clinical findings come together to identify cases that should be fully evaluated to confirm or exclude AHT. However, PredAHT should not be used as the sole piece of evidence, nor should it be considered to be a diagnostic tool that can categorically or definitively determine whether or not AHT occurred. In a legal setting, this is the job of the jury and Judge, who, like the clinician, must consider all available evidence in their decisions.

We disagree with the statement that "unless attribution is incorporated into the analysis, the data are improved or revised, and the statistical issues are resolved, arguments about AHT supported by such a model should be discounted" (Cuellar, 2017 p. 225). Regarding the quality of the data and the lack of a gold-standard diagnostic test for AHT, we fully acknowledged, and discussed, the issue of circularity in both our derivation and validation studies. Circularity is a challenge in child abuse research, and we have attempted to minimise it as far as possible. There are many diseases and diagnoses that are based upon a collection of symptoms, signs and clinical history where a gold-standard 'diagnostic test' does not exist, e.g. Kawasaki syndrome, asthma, or indeed the majority of mental health

conditions. The process of identifying such features and formulating a probability of an illness or disease, to then seek further information from additional investigations etc. is fundamental to the diagnostic process where diagnostic decisions must be made based on clinical criteria and the exclusion of differential conditions for the benefit of the patient. One simply cannot make any diagnosis without including an assessment of the physical findings. As stated above, the tool is not to be used alone but in combination with the clinical and psychosocial history, following consideration of other clinical findings and differential diagnoses. We do not believe that evidence regarding the diagnosis of AHT should be simply "discounted", however imperfect.

We would agree with Cuellar that a (rounded) probability score of 100% gives an uncomfortable level of diagnostic certainty (Maguire et al., 2011, Figure 2), however these were the numbers that arose from the data included within the original derivation model. We stress that Figure 2 (Maguire et al., 2011), which provides predicted probabilities for the 64 possible combinations of features, is a guide for clinicians, and should only be used as such. Specifically, we state that the tool "may provide a statistical estimate to assist clinicians" (Maguire et al., 2011). We have since calculated likelihood ratios for each combination of the clinical features in our model, as a formal way of incorporating other information regarding each child via the prior probability of AHT. At the time we believed that percentages were easier to interpret, and more accessible in a clinical setting.

Regarding the modelling process used to develop PredAHT, we used a well-established process to derive and compare models, and chose to exclude non-significant variables in the interests of face validity, at the expense of a possibly superior predictive model. A statistical model will only ever be an approximation to the underlying reality (Steyerberg, 2009). Cuellar rightly notes that the results of the full, and reduced models, did not differ dramatically, but we reiterate that the model has been externally validated using a

novel data set other than the one from which it was derived, and that it performed reasonably well in this validation (Cowley et al., 2015).

We fully acknowledged the large amount of missing data for some of our variables, which is unavoidable in this area as some of the investigations may not have been clinically indicated, and we chose a recognised methodology to minimise the impact of this by imputing the data. The authors criticism that long-bone fractures are "likely to be missing because it was absent rather than because the doctor did not know or forgot to write it down" (Cuellar, 2017 p. 227) misses the nuances of the clinical context and assessment process in suspected AHT cases. It is well recognised by clinicians that fractures, particularly classic metaphyseal lesions, infant rib fractures, or healing fractures, all of which are highly significant in assessing possible physical abuse, can be clinically occult and thus not identified by clinicians without radiological investigations. Ironically, that the data for this feature were missing for 100% of children in one study means that these data were likely missing completely at random, which automatically justifies the missing at random assumption and therefore the validity of our imputation approach in this particular case.

We are confused by the criticisms that the sample of patients and sample of physicians were collected in a non-randomized manner, that our sample is biased and that we therefore cannot generalize our findings to all children in the UK. An important feature of clinical prediction tools is that the dataset on which they are based should be representative of their target population (Lee et al., 2016), something Cuellar later states herself (Cuellar, 2017 p. 230). External validation studies are then conducted in order to ensure generalizability to other settings (Toll et al., 2008). Indeed, our data were not collected in a randomized manner. The datasets were acquired following a rigorous systematic review that identified the highest quality published studies. Each dataset was comprised of consecutive, population-based cases of all children presenting with head injury (Maguire et al., 2009). A wide range of clinicians

evaluated these children according to national standards and guidelines and standardized study protocols, in three countries. To address the point that these data do not reflect the whole of the UK, one dataset was based on a national reporting system (Hobbs et al., 2005) and one dataset was based on regional case ascertainment (Kemp et al., 2003), and were as representative as it would be possible to achieve. We believe that our rigorous processes enabled us to select the highest quality data for analysis. In addition, we used random effects models to account for between-study heterogeneity; this allows us to generalize to a notional wider population of quality studies of AHT.

Finally, we agree with the author that our research question is one of "backcasting", and have never purported to present it otherwise, and never purported to determine a causal relationship. We estimated the probability of AHT given the injuries seen, with no premise that any injuries seen occurred at a given time (or even the same time) or by a given mechanism. We believe that this is a valid approach to aid the identification of children who may have suffered AHT. Regardless of their timing or individual aetiology, certain combinations of clinical features at the time of an intracranial injury in a child less than three years of age yield a high probability of abuse. We also contend that, although a causal approach may be necessary in a court setting to identify a perpetrator, PredAHT is not intended for this nor is it capable of doing this, but rather it provides the clinician with a valid analysis of the probability of AHT when a specific constellation of features are present such that the clinician should be alert to the likelihood of the condition and investigate the case further. The six features within the PredAHT are 'broad brush' clinical features, the detail and nuanced characteristics of which, together with other recognised clinical aspects of head injury in young children, enable the clinician to make more precise and clinically informed decisions when confirming or excluding a diagnosis of AHT.

We believe it is important that the clarifications provided above are available to the readership of the original article, in order to prevent readers misinterpreting literature that has an important part to play in ensuring the safety of children.

COWLEY, L. E., MORRIS, C. B., MAGUIRE, S. A., FAREWELL, D. M., and KEMP, A. M. (2015), "Validation of a Prediction Tool for Abusive Head Trauma," *Pediatrics*, **136**, 290–298.

CUELLAR, M. (2017), "Causal Reasoning and Data Analysis: Problems with the Abusive Head Trauma Diagnosis", *Law, Probability and Risk*, **16**, 223–239.

HOBBS, C., CHILDS, A-M., WYNNE, J., LIVINGSTON, J., and SEAL, A. (2005). "Subdural haematoma and effusion in infancy: an epidemiological study", *Archives of Disease in Childhood,* **90**, 952–955.

KEMP, A.M., STOODLEY, N., COBLEY, C., COLES, L., and KEMP, K.W. (2003). "Apnoea and Brain Swelling in Non-Accidental Head Injury", *Archives of Disease in Childhood,* **88**, 472–476.

LEE, Y., HEEJUNG BANG, H., and KIM, D.J. (2016), "How to Establish Clinical Prediction Models," *Endocrinology & Metabolism (Seoul)*, **31**, 38–44.

MAGUIRE, S. A., KEMP, A. M., LUMB, R. C., and FAREWELL, D. M. (2011), "Estimating the Probability of Abusive Head Trauma: A Pooled Analysis," *Pediatrics*, **128**, e550–e564.

MAGUIRE, S., PICKERD, N., FAREWELL, D., MANN, M., TEMPEST, V., and KEMP, A. (2009), "Which Clinical Features Distinguish Inflicted from Non-Inflicted Brain Injury? A Systematic Review," *Archives of Disease in Childhood*, **94**, 860–867.

STEYEBERG, E. (2009), *"Clinical Prediction Models: A Practical Approach to Development, Validation and Updating,"* New York: Springer.

TOLL, D.B., JANSSEN, K.J., VERGOUWE, Y., and MOONS, K.G. (2008), "Validation, Updating and Impact of Clinical Prediction Rules: A Review". *Journal of Clinical Epidemiology*, **61**, 1085–1094.